Distributed Data fusion for detecting Sybil attacks in VANETs

Nicole El Zoghby, Véronique Cherfaoui, Bertrand Ducourthial, Thierry Denoeux

▶ **To cite this version:**

Nicole El Zoghby, Véronique Cherfaoui, Bertrand Ducourthial, Thierry Denoeux. Distributed Data fusion for detecting Sybil attacks in VANETs. 2nd International Conference on Belief Functions (BELIEF 2012), May 2012, Compiègne, France. pp.351-358, 10.1007/978-3-642-29461-7_41 . hal-00756330

HAL Id: hal-00756330
https://hal.science/hal-00756330

Submitted on 22 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed Data fusion for detecting Sybil attacks in VANETs

Nicole El Zoghby, Véronique Cherfaoui, Bertrand Ducourthial, Thierry Denœux

Abstract Sybil attacks have become a serious threat as they can affect the functionality of VANETs (Vehicular Ad Hoc Networks). This paper presents a method for detecting such attacks in VANETs based on distributed data fusion. An algorithm has been developed in order to build distributed confidence over the network under the belief function framework. Our approach has been validated by simulation.

1 Introduction

Exchanging data in a Mobile Ad hoc NETwork (MANET) in a safe manner becomes an important issue. These networks are vulnerable to different attacks such as intrusion. The need for security requires the introduction of the notion of confidence, as each node should have confidence in other nodes or in the received data before using the exchanged information in different applications. By broadcasting messages, nodes will discover their neighborhood. These neighbors can be fake or real nodes, they can also be attackers. Different research papers have been dedicated to find a solution to these problems. Many recent works deal with reputation mechanisms ([20],[1],[9]) and trust evaluation ([16],[17]) to manage the confidence in the source of information. Others were interested in data aggregation without taking into account the source [2][3][10][13].

We propose a method to fuse data in a distributed system in order to build confidence over the network. Nodes broadcast their opinions, which are then used at the reception to evaluate other nodes. Since local opinion is uncertain and incomplete, the use of belief functions to evaluate the received messages seems appropriate. The fusion of a node's local knowledge with all the received messages is done by Demp-

Nicole El Zoghby, Véronique Cherfaoui, Bertrand Ducourthial, Thierry Denœux
Heudiasyc UMR CNRS 7253, Université de Technologie de Compiègne, France, e-mail: nicole.el-zoghby@hds.utc.fr , veronique.cherfaoui@hds.utc.fr , bertrand.ducourthial@hds.utc.fr , thierry.denoeux@hds.utc.fr

ster's rule. The network can suffer from cycles of data dissemination where the same information can be combined many times as it is coming from independent sources [14],[11]. To avoid that, we use the cautious rule of combination [5].

We are interested in studying the confidence in a node for the purpose of detecting sybil attacks in VANETs (Vehicular Ad Hoc NETworks). The sybil attack is the case where a single faulty entity, called a malicious node, can present multiple identities [6] known as sybil nodes or fake nodes. This attack can affect the functionality of the network for the benefit of the attacker. Several techniques have been developed to detect misbehaving or fake nodes in VANETs. Gole et al [7] represented an adversial parsimony that means finding the explanation for corrupted data. Vehicles can distinguish their neighbors by using cameras or exchanging messages in infrared light spectrum. The technique described by Xiao et al. is based on statistic signal strength analysis with the help of roadside infrastructure to detect sybil nodes [18]. Yan et al. [19] used an on-board radar to detect neighbors and to confirm their announced position. Piro et al [12] showed that the sybil attack can be detected passively through single or multiple observers. Due to the dynamics of the vehicular networks, of the number of vehicles and of the lack of permanent infrastructure access, deploying a Public Key Infrastructure in vehicular network (Vehicular PKI) is a very challenging task. As shown in [8], by simply comparing the received signal strength, half of the vehicles can detect the Sybil nodes and it is expected that cooperative techniques would decrease the number of cheated vehicles. Our work proposes such a cooperative algorithm between vehicles, based on the theory of belief functions, and could allow to avoid cryptographic schemes.

In this paper, we develop a distributed fusion technique based on the theory of belief functions. We first describe the system and how we represent the confidence using mass functions. We present the distributed data fusion approach and the proposed algorithm. We validate our approach by simulations and finally we conclude.

2 Distributed data fusion approach

We consider a network composed by nodes exchanging messages. It can be modeled by a directed graph $G = (V,E)$, where V represents the set of nodes $V = \{v_1, v_2, ..., v_n\}$ and E represents the set of edges. The neighbors of each node are represented by $\Gamma(v) = \{v_j \in V, \{v_i, v_j\} \in E\}$. For the sake of simplicity, we suppose that each node knows $n = |V|$. Figure 1 shows an example of network configuration. Each node periodically sends *regular messages* composed of its true identity and geographical position. Moreover, one of the node sends both its regular messages and *fake messages* composed of a forged identity and a forged position. By receiving the fake messages, other nodes are cheated and consider a non existing node, called *fake node* or *Sybil node*. We consider a single malicious node, which creates several Sybil nodes. All nodes use the same transmission system (same antenna, same transmission power). The topology of the network is given by the transmission radio range of the nodes (unit disk graph). We propose a data fusion methodology to

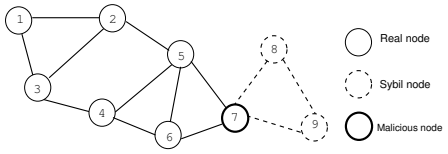

Fig. 1 Network Configuration

combine data exchanged in a mobile ad hoc network, with the aim of quantifying the confidence in the nodes of the network. For this purpose, the messages contain also the sender's confidence in the nodes of the network.

Representing the confidence by mass function: Each node is able to assign a confidence to each other node of the network. This confidence is represented by a basic belief assignment (bba) denoted by m, defined from the frame of discernment $\Omega = \{0,1\}$ where 0 represents $FakeNode$ and 1 represents $RealNode$.

We denoted by m_{ij} the corresponding *bba* that represents the opinion of node v_i about node v_j. The bba m_{ij} is defined in Ω by:

$$m_{ij}(\emptyset) = 0 \; ; \; m_{ij}(0) = p_{ij} \; ; \; m_{ij}(1) = q_{ij} \; ; \; m_{ij}(\Omega) = 1 - p_{ij} - q_{ij}. \tag{1}$$

Principle of the approach: Node v_k sends a message to v_i containing its identity, its coordinates and its opinion about the network. When node v_i receives the message, it calculates, after analyzing the signal strength, what we call a direct confidence. It is a mass vector denoted by $m_{d_{ik}}$. This direct confidence is saved in a local memory called *local knowledge* or *private knowledge*.

Note that each node has two bodies of knowledge: *local knowledge* and *public knowledge*. Local knowledge represents what each node can collect from its neighborhood. It is combined with the public knowledge of other nodes in order to update the public knowledge and rebroadcast it through the network. We thus have a distributed system. Local knowledge depends only on the signal strength of the messages and not on their content: consequently, it cannot be cheated. In contrast, public knowledge is based on the combination of the content of the messages and can be cheated by fake messages. This is why we separate local and public knowledge. The internal memory of each node is thus represented by two mass vectors (arrays of $|V|$ cells initialized at $m(\Omega)$ if $i \neq j$ and $m(1)$ if $i = j$):

$$Kprivate_i(t) = [m_{ij}^{(t)}] \; ; \quad Kpublic_i(t) = [m_{p_{ij}}^{(t)}]. \tag{2}$$

Distributed fusion Algorithm: The processing steps performed at the reception are presented in Algorithm 1, and explained hereafter.

Distributed Fusion: When node v_i receives a message, it computes the direct confidence $m_{d_{ik}}$. This confidence is independent of previous messages and it is not the result of any other combination. So we use it to update the receiver's local knowledge about the transmitter by Dempster's rule [4]. The function UpdateLocalKnowledge $(m_{l_{ik}}^{(t-1)}, m_{d_{ik}}^{(t)})$ is calculated as:

Algorithm 1: Received Message Processing on node v_i

Require: message from v_k to v_i, the signal strength P, message contains $m_{p_{kj}} \forall j$

Ensure: $Kprivate_i$=$[m_{l_{ij}}^{(t)}]$ and $Kpublic_i$=$[m_{p_{ij}}^{(t)}]$ $\forall j \in V$

$m_{d_{ik}}^{(t)} \leftarrow DirectConfidence(message, P)$

$m_{l_{ik}}^{(t)} \leftarrow UpdateLocalKnowledge(m_{l_{ik}}^{(t-1)}, m_{d_{ik}}^{(t)})$

$m_{p_{ik}}^{(t)} \leftarrow UpdatePublicKnowledge(m_{p_{ik}}^{(t-1)}, m_{l_{ik}}^{(t)})$

$\alpha \leftarrow DiscountingFactor(m_{l_{ik}})$

for *each node* $j \in V$ *such as* $j \neq i$, $j \neq k$ **do**

$\quad {}^{\alpha}m_{p_{kj}}^{(t)} \leftarrow DiscountTransmitterKnowledge(\alpha, m_{pkj}{}^{(t)}, m_{\Omega}^{(t)})$

$\quad m_{p_{ij}}^{(t)} \leftarrow UpdatePublicKnowledge(m_{p_{ij}}^{(t-1)}, {}^{\alpha}m_{p_{kj}}^{(t)})$

$$m_{l_{ik}}^{(t)} = m_{l_{ik}}^{(t-1)} \oplus m_{d_{ik}}^{(t)}, \tag{3}$$

where $\oplus$ denotes Dempster's rule. Since fake nodes might falsify the opinion of each node, the knowledge of other nodes is needed. To this end we use a distributed fusion to collect other opinions. As we consider that the transmitter is not totally reliable, we discount its opinion before combining it with the receiver's knowledge. The discounting factor $\alpha = 1 - m_{l_{ik}}(1)$ is defined as the plausibility that the transmitter is unreliable. The transmitter's opinion is discounted with the function DiscountTransmitterKnowledge$(\alpha, m_{kj}^{(t)}, m_{\Omega}^{(t)})$ as follows:

$${}^{\alpha}m_{p_{kj}}^{(t)} = (1-\alpha).m_{p_{kj}}^{(t)} + \alpha.m_{\Omega}^{(t)}. \tag{4}$$

To update the receiver's public knowledge, we use the cautious rule [5]. In a distributed system, the same information can be received and treated many times. While combining the knowledge, it is useful to use an idempotent rule to avoid counting the same information several times (data incest) as if it is provided by different independent sources. So the function UpdatePublicKnowledge$(m_{p_{ij}}^{(t-1)}, {}^{\alpha}m_{p_{kj}}^{(t)})$ allows us to combine the receiver's public knowledge with the transmitter's discounted knowledge about its neighbors as follows:

$$m_{p_{ij}}^{(t)} = m_{p_{ij}}^{(t-1)} \circledwedge {}^{\alpha}m_{p_{kj}}^{(t)}, \tag{5}$$

where $\circledwedge$ denotes the cautious rule.

Direct confidence: Different methods can be used to compute the direct confidence $m_{d_{ik}}$. We propose a method that allows us to convert a real measure into a mass function. The real measure is based on signal strength analysis. Each receiver can analyze the signal strength to detect if the announced position is the real one [8]. It measures the strength of the received signal and calculates a theoretical value in terms of the node's coordinates. The estimated value of the signal strength is calculated by the Friis formula as $\mu = P_e.G_{SR}/d_{ik}^2$, where

- P_e is the transmitted signal power, depending on the transmitter antenna;
- $G_{SR} = \frac{G_t.G_r.\lambda^2}{16.\pi^2}$ is the antenna gain, G_t and G_r are the gains of the transmit antenna and the receive antenna, respectively, and λ is the wavelength;
- d_{ik} is the distance between the transmitter node v_k and the receiver node v_i.

The comparison between the estimated power and the theoretical one allows the detection of a misbehavior. We propose to compute the plausibility that the received signal power P is equal to x, given that the transmitting node is a true node ($\omega = 1$) as follows:

$$pl(P = x/\omega = 1) = \frac{f(x/\omega=1)}{\sup_{x' \in \mathbb{R}}(f(x'/\omega=1)} , \tag{6}$$

where $f(x/\omega = 1)$ is the normal density function with mean μ and standard deviation σ depending on the receiver antenna.

The plausibility $pl(P = x/w = 0)$ is defined as shown in Figure 2: if the estimated and the theoretical powers are equal, we leave the possibility that the transmitter can be a fake node. Indeed, if the transmitter is a fake node but its position is near the malicious node therefore the estimated position will be approximately equal to the measured position. This result can influence the detection of the fake node.

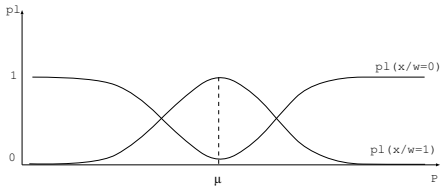

Fig. 2 Plausibility of received power values for true ($\omega = 1$) and fake ($\omega = 0$) nodes

The direct confidence is computed using the Generalized Bayes theorem [15]. It is obtained by combining the prior knowledge about the transmitter m_0^{Ω} with the plausibility that the node is a fake node knowing that it is a real $\{0\}^{pl(x/w=1)}$ and the plausibility that the node is a real node knowing that it is a fake $\{1\}^{pl(x/w=0)}$:

$$m_{d_{ik}}^{(t)} = m^{\Omega}(./x) = m_0^{\Omega} \textcircled{\cap} \{0\}^{pl(x/w=1)} \textcircled{\cap} \{1\}^{pl(x/w=0)}, \tag{7}$$

where $\textcircled{\cap}$ denotes the unnormalized Dempster's rule.

3 Results

In order to validate our approach, Algorithm 1 has been implemented in Matlab. Simulations were performed on static and dynamic network. For simplicity of analysis, we first assumed all nodes in the network to be static. We performed simulations on different random network configurations. Next we tested our approach on

a dynamic network, where nodes were moving in the same direction following a highway scenario.

Implementation: In this part we will represent an example of a network composed from six true nodes, one of them is a malicious node that creates three fake nodes. The transmitted signal Power P_e is about 600 mW and the antenna ranges is in order of 400 m. We consider that each transmitter sends its *id*, its *position* and its *public knowledge*. The receiver uses these informations to perform all the calculations and to verify if the node is true or fake. Simulations are performed until the convergence of the algorithm. We consider that the algorithm has converged when $|m_{ij}^{(t-1)} - m_{ij}^{(t)}| < \varepsilon$, where ε is a defined small threshold. The results of the simulation will be represented by gray scale matrices.

Static network: We present in Figure 3 an example of a network configuration where the nodes are static (left figure) and the result of the simulation (right figure). The white color in the right figure corresponds to a mass equal to 1 representing true nodes. The black color correspond to a mass equal to 0 representing fake nodes. The malicious node 3 will try to convince other nodes that the fake nodes (7,8,9) are true nodes. The fake nodes have the same opinion as the malicious node. The first part of the rightmost figure represents the private knowledge. Each node has only information about its neighbors. The second part represents the public knowledge. We see that $m_{p_{ij}}(\{1\}) = 0$ for $i = \{1,2,4,5,6\}$ and $j = \{7,8,9\}$, which means that the true nodes have correctly identified nodes $\{7,8,9\}$ as untrustworthy.

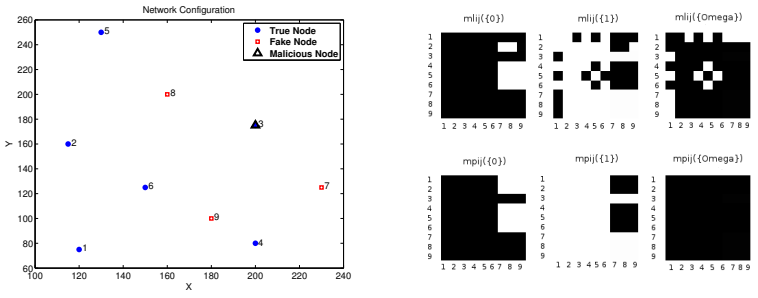

Fig. 3 Network configuration and simulation results.

To verify the convergence of the algorithm, we performed simulations on different random network configurations by changing the number of the fakes nodes. Table 1 represents the result with different proportions of fake nodes. Each iteration represents the simulation of the process of a message. It needs more time to converge when the proportion of the fake nodes is greater. Our approach can detect sybil nodes with different static configurations.

Dynamic Network: Static configurations have some limits, especially when a malicious node is not in the neighborhood of the true nodes: in that case, fake nodes cannot be detected. So, we simulated a dynamic scenario in which nodes move in

Table 1 Results with different nodes configurations

Nodes Configurations	Average of the number of iterations[a]	Standard deviation
True Nodes=6 Fake Nodes =3	207.05	7.86
True Nodes=6 Fake Nodes =4	227.55	6.89
True Nodes=6 Fake Nodes =5	255.8	6.33
True Nodes=6 Fake Nodes =6	304.7	7.55

[a] These results represent the average of 20 simulations.

the same direction as on a highway. While moving, the neighborhood of each node changes. It influences the private knowledge because it depends on the neighborhood. Thanks to public knowledge, each node can get information about the whole network and can quantify its confidence. Table 2 shows results for different dynamic network configurations. The number of iterations until convergence changes at each simulation, because the node motions and neighborhoods are random. These preliminary results suggest that true nodes can successfully detect fake nodes in the network while moving on a highway.

Table 2 Results for dynamic networks with different node configurations

Nodes Configurations	Average of the number of iterations[a]	Standard deviation
True Nodes=6 Fake Nodes =3	119.3	45.88
True Nodes=6 Fake Nodes =4	274.4	40.96
True Nodes=6 Fake Nodes =5	361.1	54.23
True Nodes=6 Fake Nodes =6	376.3	32.05

[a] These results represent the average of 10 simulations.

4 Conclusion

A distributed data fusion approach based on belief functions for detecting sybil attacks in VANETs has been developed. The method uses both Dempster's rule and cautious rule to combine information and to compute a distributed confidence over the network. The results are promising and demonstrate that we can determine the reliability of nodes and detect fake nodes in a VANET. More realistic scenarios are currently being studied using an ad hoc network simulator.
The method presented in this paper computes the confidence in the nodes without taking into account the contents of the messages exchanged in the network. The joint analysis of information and node reliability is currently being investigated. Results along these lines will be reported in future publications.

References

1. M.P. Singh B. Yu. An evidential model of distributed reputation management. In *First international Joint Conference on Autonomous Agents and Multi-Agents Systems, ACM Press*, pages 294–301, Bologna, Italy, 2002.
2. T. M. Chen and V. Venkataramanan. Dempster-shafer theory for intrusion detection in ad hoc networks. In *IEEE Internet Computing*, volume 9, pages 35–41, 2005.
3. V. Cherfaoui, T. Denoeux, and Z. L. Cherfi. Distributed data fusion: application to confidence management in vehicular networks. In *11th Int. Conf. on Information Fusion*, pages 846–853, Germany, 2008.
4. A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
5. T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172:234–264, 2008.
6. J.R Douceur. The sybil attack. In *the International Workshop on Peer to Peer Systems*, pages 251–260, Cambridge, MA, USA, 2002.
7. P. Golle, D. Greene, and J. Staddon. Detecting and correcting malicious data in vanets. In *1st ACM Workshop on Vehicular Ad hoc Networks (VANET)*, pages 29–37, New York, NY, USA, 2004.
8. G. Guette and B. Ducourthial. On the sybil attack detection in vanet. In *International Workshop on Mobile Vehicular Networks (MoveNet 2007), co-located with IEEE MASS 2007*, Pisa, October 2007.
9. J. Liu and V. Issarny. Enhanced reputation mechanism for mobile ad hoc networks. In *2nd International Conference on Trust Management*, pages 48–62, Oxford, UK, 2004.
10. C. Lochert, B. Scheuermann, and M. Mauve. Probabilistic aggregation for data dissemination in vanets. In *4th ACM international Workshop on Vehicular Ad Hoc Networks*, pages 1–8, Montral, QC, Canada, 2007.
11. H.B. Mitchell. *Multisensor Data Fusion: An introduction*. Springer, 2007.
12. C. Piro, C. Shields, and B.N Levine. Detecting the sybil attack in mobile ad hoc networks. In *IEEE/ACM Intl Conf on Security and privacy in Communication Networks (SecureComm)*, pages 1–11, August 2006.
13. M. Raya, P. Papadimitratos, V. D. Gligor, and J p. Hubaux. On data-centric trust establishment in ephemeral ad hoc networks. In *the 28th IEEE conference on Computer Communications (INFOCOM)*, pages 1238–1246, Phoenix, AZ., USA, April 2008.
14. R. J. Evans S. Mclaughlin, V. Krishnamurthy. Bayesian network model for data incest in a distributed sensor network. In *the 7 th International Conference on Information Fusion*, volume 1, Stockholm, Sweden, 2004.
15. Ph. Smets. Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9:1–35, 1993.
16. G. Theodorakopoulos and J. S. Baras. Trust evaluation in ad-hoc networks. In *ACM Workshop Wireless Security*, pages 1–10, Philadelphia, PA, USA, 2004.
17. J. Wang and H j. Sun. A new evidential trust model for open communities. *Computer Standards & Interfaces*, 31:994–1001, 2009.
18. B. Xiao, B.Yu, and C.Gao. Detection and localization of sybil nodes in vanets. In *the Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks*, pages 1–8, Los Angeles, CA,USA, 2006.
19. G. Yan, G. Choudhary, M. Weigle, and S. Olariu. Providing vanet security through active position detection. *Computer Communications: Special Issue on Mobility Protocols for ITS/ VANET*, 31(12):2883–2897, 2008.
20. G. Zacharia and P. Maes. Trust management through reputation mechanisms. *Applied Artificial Intelligence*, 14:881–907, 2000.